**Assignment: Data Science Masters**

**Q1: In order to predict house price based on several characteristics, such as location, square footage, number of bedrooms, etc., you are developing an SVM regression model. Which regression metric in this situation would be the best to employ?**

* **Answer**: The best metric to employ depends on the specific goal:
  + If the focus is on how well the model explains the variance in the data, use **R-squared**.
  + If the focus is on minimizing the prediction error, use **Mean Absolute Error (MAE)** or **Root Mean Squared Error (RMSE)**.
  + **RMSE** is generally preferred for house price predictions, as it penalizes larger errors more than MAE.

**Q2: You have built an SVM regression model and are trying to decide between using MSE or R-squared as your evaluation metric. Which metric would be more appropriate if your goal is to predict the actual price of a house as accurately as possible?**

* **Answer**: **Mean Squared Error (MSE)** would be more appropriate as it directly measures the average squared difference between the predicted and actual values, making it better suited for assessing prediction accuracy.

**Q3: You have a dataset with a significant number of outliers and are trying to select an appropriate regression metric to use with your SVM model. Which metric would be the most appropriate in this scenario?**

* **Answer**: **Mean Absolute Error (MAE)** is the most appropriate metric in this scenario because it is less sensitive to outliers compared to MSE or RMSE.

**Q4: You have built an SVM regression model using a polynomial kernel and are trying to select the best metric to evaluate its performance. You have calculated both MSE and RMSE and found that both values are very close. Which metric should you choose to use in this case?**

* **Answer**: **Root Mean Squared Error (RMSE)** should be chosen because it is on the same scale as the target variable, making it more interpretable than MSE.

**Q5: You are comparing the performance of different SVM regression models using different kernels (linear, polynomial, and RBF) and are trying to select the best evaluation metric. Which metric would be most appropriate if your goal is to measure how well the model explains the variance in the target variable?**

* **Answer**: **R-squared** is the most appropriate metric for measuring how well the model explains the variance in the target variable, as it provides a normalized measure of fit.